CH5350: Applied Time-Series Analysis

Random Variables and Moments: Review

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

Random Variable

Informal definition

A **random variable** (RV) is one whose value set contains at least two elements, i.e., it draws one value from at least two possibilities. The space of possible values is known as the **outcome space or sample space**.

Formal definition (Priestley, 1981)

A random variable X is a mapping from the sample space $\mathbb{S}$ onto the real line s.t. to each element $s \subset \mathbb{S}$ there corresponds a unique real number.

Focus of this course: **Continuous-valued random variables**

Distribution and Density functions

Probability distribution function

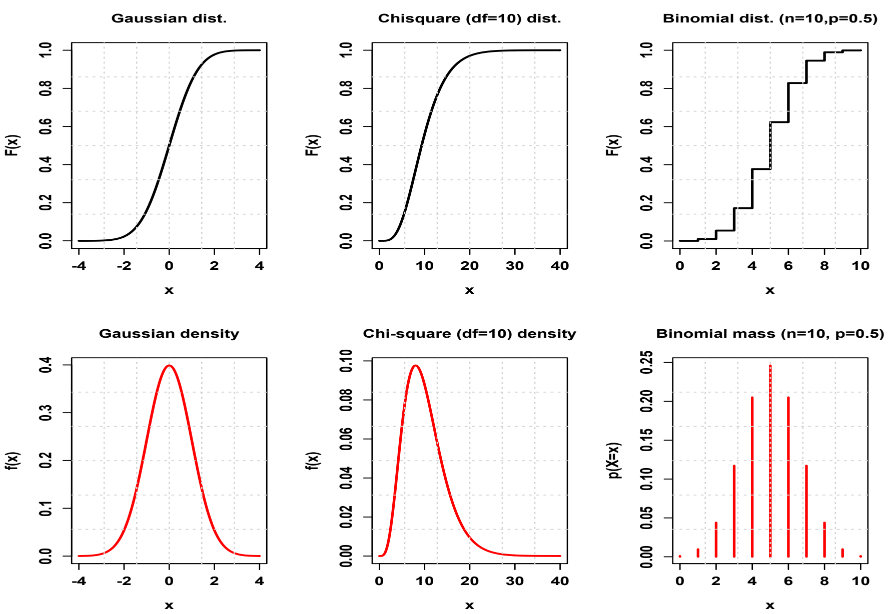
Also known as the **cumulative distribution function**: $\boxed{F(x) = \Pr(X \leq x)}$

When the density function exists, i.e., for continuous-valued RVs,

Density functions

1. Area yields probability $\boxed{\Pr(a < x < b) = \int_a^b f(x)\, dx} \quad \Longrightarrow \int_{-\infty}^{\infty} f(x)\, dx = 1$
2. Derivative of c.d.f.: $\boxed{f(x) = \dfrac{dF(x)}{dx}}$

Examples: c.d.f. and p.d.f.



The type of distribution for a random phenomenon depends on its nature.

Practical Aspects: Moments of a p.d.f.

- It may not be necessary to know the p.d.f. in practice!
- What is of interest in practice is (i) **the most likely value and/or the expected outcome** (mean) and (ii) **how far the outcomes are spread** (variance). These are related to what are known as moments of the density function.

The n^{th} moment of a p.d.f. is defined as

$$\boxed{M_n(X) = \int_{-\infty}^{\infty} x^n f(x)\, dx} \tag{1}$$

Linear random process and moments

It turns out that for linear processes, predictions of random signals and estimation of model parameters it is sufficient to have the knowledge of **mean**, **variance** and **covariance** (for the multivariate case), *i.e.*, it is sufficient to know the first and second-order moments of p.d.f.

Mean and Variance: Important moments!

Mean

Mean, a.k.a. the **expectation** of the RV, is defined as

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f(x)\, dx \tag{2}$$

Variance

The variance of a RV measures the average spread of outcomes around its mean,

$$\sigma_X^2 = E((X-\mu_X)^2) = \int_{-\infty}^{\infty} (x-\mu_X)^2 f(x)\, dx \tag{3}$$

Note: Integration in (2) and (3) are across the **outcome space** and NOT across time.

Remarks

- Prediction perspective:

  The mean is the best prediction $\hat{X}$ of the RV X in the minimum mean square error sense, *i.e.*,

  $$\mu = \arg\left[\min_c E(X-\hat{X})^2 \text{ s.t. } \hat{X} = c\right]$$

- Applying the **expectation operator** E to a RV or its function produces its "average" or expected value.

$$\boxed{E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx} \tag{4}$$

Points to note

- As (3) suggests, σ_X^2 is the **second central moment** of $f(x)$. Further,

$$\sigma_X^2 = E(X^2) - \mu_X^2 \tag{5}$$

- The variance definition is in the space of outcomes. **It should not be confused with the widely used variance definition for a series or a signal (sample variance)**.
- Large variance indicates far spread of outcomes around its statistical center. Naturally, in the limit as $\sigma_X^2 \to 0$, X becomes a **deterministic** variable.

Bivariate / Multivariate Analysis

Joint density

Where bivariate analysis is concerned, we start to think of **joint probability density functions**.

Consider two continuous-valued RVs X and Y. The probability that these variables take on values in a rectangular cell is given by the **joint density**

$$Pr(x_1 \leq x \leq x_2, y_1 \leq y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x,y)\, dx\, dy$$

Marginal and conditional p.d.f.s

Associated with this joint probability (density function), we can ask two questions:

1. What is the probability $\text{Pr}(x_1 \leq X \leq x_2)$ regardless of the outcome of Y and vice versa? (**marginal density**)
2. What is the probability $\text{Pr}(x_1 \leq X \leq x_2)$ given Y has occurred and taken on a value $Y = y$? (**conditional density**)
   - Strictly speaking, one cannot talk of Y taking on an exact value, but only of values within an infinitesimal neighbourhood of y.

Marginal and Conditional densities

Marginal density

The **marginal density** of a RV X (Y) with respect to another RV Y (X) is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \qquad \text{likewise,} \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx \tag{6}$$

Conditional density

The conditional density of Y given $X = x$ (strictly, between x and $x + dx$) is

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f(x)} \tag{7}$$

Conditional Expectation and Prediction

In several situations we are interested in "predicting" the outcome of one phenomenon given the outcome of another phenomenon.

Conditional expectation

The conditional expectation of Y given $X = x$ is

$$E(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X=x}(y)\,dy = \phi(x) \tag{8}$$

The conditional expectation is the best predictor of Y given X among all the predictors that minimize the mean square prediction error

Independence

(Frequentist school)

Two random variables are said to be **independent** if and only if

$$\boxed{f(x,y) = f_X(x) f_Y(y)} \tag{9}$$

Alternatively (Bayesian school)

Two random variables are said to be **independent** if and only if

$$f_Y(y|x) = f_Y(y) \qquad \text{or} \qquad f_X(x|y) = f_X(x) \tag{10}$$

Covariance

The statistic that measures the **co-variation** of two RVs is given by the second-order moment of the joint p.d.f.:

$$\sigma_{XY} = E((X-\mu_X)(Y-\mu_Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_X)(y-\mu_Y) f(x,y)\, dx\, dy \tag{11}$$

$$= E(XY) - E(X)E(Y) \tag{12}$$

It is useful to collect the variations and co-variation in a single **variance-covariance** matrix

$$\Sigma_X = \begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} \\ \sigma_{X_2X_1} & \sigma^2_{X_2} \end{bmatrix}$$

Covariance for vector quantities

In the general scenario, for a vector of random variables,

$$\mathbf{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix}^T$$

we work with the variance-covariance matrix

$$\Sigma_{\mathbf{X}} = E((\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^T) = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_N} \\ \sigma_{X_2X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2X_N} \\ \vdots & \cdots & \cdots & \vdots \\ \sigma_{X_NX_1} & \sigma_{X_NX_2} & \cdots & \sigma_{X_N}^2 \end{bmatrix}$$

Properties of the covariance matrix

- It is a **symmetric** measure: $\sigma_{XY} = \sigma_{YX}$, i.e., it is not a directional measure. **Consequently it cannot be used to sense causality** (cause-effect relation).
- The covariance matrix $\Sigma_{\mathbf{X}}$ is a **symmetric, positive semi-definite** matrix $\implies \lambda_i(\Sigma_{\mathbf{X}}) \geq 0 \ \ \forall\, i$
- Most importantly, **covariance is only a measure of linear relationship** between two RVs, i.e.,

  When $\sigma_{XY} = \sigma_{YX} = 0$, there is no linear relationship between X and Y
- Linear transformation of the random variables $\mathbf{Z} = \mathbf{A}\mathbf{X}$ results in

$$\Sigma_{\mathbf{Z}} = E((\mathbf{Z} - \mu_{\mathbf{Z}})(\mathbf{Z} - \mu_{\mathbf{Z}})^T) = \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T \qquad (13)$$

Correlation

To overcome the unbounded and scaling-sensitivity issues, a normalized version of covariance known as **correlation** is introduced:

$$\boxed{\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}} \tag{14}$$

- The correlation defined above is also known as Pearson's correlation
- Other forms of correlations exist, namely, **reflection correlation coefficient, Spearman's rank correlation coefficient, Kendall's tau rank correlation coefficient**

Unity correlation

Boundedness

For all bivariate distributions with finite second order moments,

$$|\rho_{XY}| \leq 1 \tag{15}$$

with equality if, with probability 1, there is a linear relationship between X and Y.

Correlation measures **linear dependence.** Specifically,

i. $|\rho_{XY}| = 1 \iff Y = \alpha X + \beta$ (Y and X are linearly related with or without an intercept)

ii. $\rho_{XY} = 0 \iff X$ and Y have no linear relationship (non-linear relationship cannot be detected)

Uncorrelated variables

Two RVs are said to be **uncorrelated** if $\sigma_{XY} = 0 \Longrightarrow \rho_{XY} = 0$. This also implies,

$$\boxed{E(XY) = E(X)E(Y)} \tag{16}$$

- Uncorrelatedness $\Longleftrightarrow$ NO **linear** relationship between X and Y.
- $|\rho_{XY}| < 1$ if and only if $Y = \alpha X + \varepsilon$, where ε is another RV due to other effects (e.g., noise, non-linearities). In fact, (assuming $\sigma_{\varepsilon X} = 0$),

$$\boxed{\rho_{YX} = \pm \frac{1}{\sqrt{1 + \dfrac{\sigma_\epsilon^2}{\alpha^2 \sigma_X^2}}} \leq \pm 1}$$

Independence vs. Uncorrelated variables

Independence $\Longrightarrow$ Uncorrelated condition but NOT vice versa.
Thus independence is a stronger condition.

If the variables X and Y have a bivariate Gaussian distribution,
Independence $\Longleftrightarrow$ Uncorrelated condition.
Therefore, in all such cases, independence and lack of correlation are equivalent.

Connections b/w Correlation and Linear regression

Correlation between two RVs is naturally related to the linear regression problem.

Given two (zero-mean) RVs X and Y, consider two related linear regression (prediction) problems.

$$\hat{Y} = bX \qquad\qquad \hat{X} = \tilde{b}Y \tag{17}$$

Denote the optimal estimates that minimize $E(\varepsilon^2) = E(Y - \hat{Y})^2$ and $E(\varepsilon^2) = E(X - \hat{X})^2$ as $b^\star$ and $\tilde{b}^\star$, respectively. Then,

$$\boxed{\rho^2_{XY} = b^\star \tilde{b}^\star} \qquad\qquad \textbf{Note:} \quad b^\star = \frac{\sigma_{XY}}{\sigma^2_X}; \quad \tilde{b}^\star = \frac{\sigma_{XY}}{\sigma^2_Y} \tag{18}$$

Zero correlation implies no linear fit in either direction.

Remarks, limitations, . . .

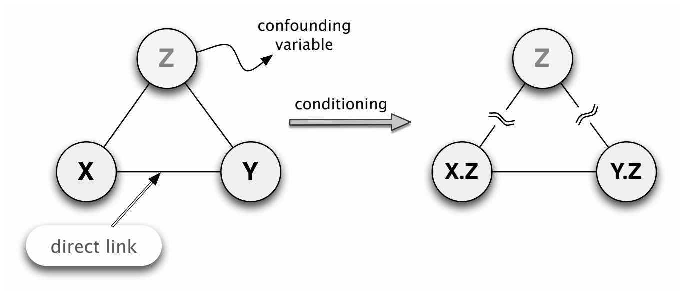
- **Correlation is only a mathematical / statistical measure.** It does not take into account any physics of the process that relates X to Y.
- High values of correlation only means that a linear model can be fit between X and Y. It does not mean that in reality there exists a linear process that relates X to Y
- Correlation is symmetric, i.e., $\rho_{XY} = \rho_{YX}$. Therefore, it is not a cause-effect measure, meaning it cannot detect direction of relationships
- Correlation is primarily used to determine if a linear model can explain the relationship:

Remarks, limitations, ... contd.

- High values of **estimated** correlation may imply linear relationship over the experimental conditions, but a non-linear relationship over a wider range.
- **Absence of correlation only implies that no linear model can be fit.** Even if the true relationship is linear, noise can be high, thus limiting the ability of correlation to detect linearity
- **Correlation measures only linear dependencies**. Zero correlation means $E(XY) = E(X)E(Y)$. In contrast, independence implies $f(x, y) = f(x)f(y)$

Despite its limitations, correlation and its variants remain one of the most widely used measures in data analysis

Resolving confounding: Conditional (partial) correlation



- If the **conditional correlation** vanishes, then the connection is purely indirect, else there exists a direct relation

Correlation measures total (linear) connectivity, whereas conditional or partial version measures "direct" association.

Partial covariance

The conditional or partial covariance is defined as

$$\sigma_{XY.Z} = \text{cov}(\epsilon_{X.Z}, \epsilon_{Y.Z}) \qquad \text{where } \epsilon_{X.Z} = X - \hat{X}^\star(Z),\ \epsilon_{Y.Z} = Y - \hat{Y}^\star(Z)$$

where $\hat{X}^\star(Z)$ and $\hat{Y}^\star(Z)$ are the **optimal predictions** of X and Y using Z.

Partial correlation (PC)

$$\boxed{\rho_{XY.Z} = \frac{\sigma_{XY.Z}}{\sigma_{\epsilon_{X.Z}}\sigma_{\epsilon_{Y.Z}}} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{(1-\rho_{XZ}^2)}\sqrt{(1-\rho_{ZY}^2)}}} \tag{19}$$

- **Partial correlation** $\equiv$ Analysis in the **inverse domain.**

Partial correlation: Example

Consider two RVs $X = 2Z + 3W$ and $Y = Z + V$ where V, W and Z are zero-mean RVs. Further, it is known that *i.e.*, $\sigma_{VW} = 0 = \sigma_{VZ} = \sigma_{WZ}$.
Evaluating the covariance between X and Y yields

$$\sigma_{YX} = E((2Z + 3W)(Z + V)) = 2E(Z^2) = 2\sigma_Z^2 \neq 0$$

implying X **and** Y **are correlated.** However, applying (19), it is easy (with a bit of effort) to see that

$$\rho_{YX.Z} = 0$$

Thus, X **and** Y **are not correlated given** Z**.**

Partial correlation by inversion

Partial correlation coefficients can also be computed from the inverse of **covariance** (or **correlation**) matrix as follows.

Assume we have M RVs $X_1, X_2, \cdots, X_M$. Let $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_M \end{bmatrix}^T$.

1. Construct the covariance (or correlation) matrix $\Sigma_{\mathbf{X}}$ (or $\Omega_{\mathbf{X}}$).
2. Determine the inverse of covariance (or correlation) matrix, $\mathbf{S_X} = \Sigma_{\mathbf{X}}^{-1}$ (or $\mathbf{P_X} = \Omega_{\mathbf{X}}^{-1}$).
3. The partial correlation between X_i and X_j conditioned on all $\mathbf{X} \backslash \{X_i, X_j\}$ is then:

$$\rho_{X_i X_j . \mathbf{X} \backslash \{X_i, X_j\}} = -\frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}} = -\frac{p_{ij}}{\sqrt{p_{ii}}\sqrt{p_{jj}}} \tag{20}$$

Partial correlation and Linear regression

Partial correlation enjoys a relation with the regression models as does correlation.

Consider three random variables X_1, X_2, X_3. To understand the PC between X_1 and X_2 conditioned on X_3, construct the "forward" and "reverse" linear predictors:

$$\hat{X}_1 = b_{12}X_2 + b_{13}X_3, \qquad \hat{X}_2 = b_{21}X_1 + b_{23}X_3 \tag{21}$$

Then, we have an interesting result:

The squared-PC between X_1 and X_2 is the product of optimal coefficients

$$\boxed{\rho_{12.3}^2 = b_{12}^\star b_{21}^\star}, \qquad b_{12}^\star = \frac{\sigma_1}{\sigma_2}\left(\frac{\rho_{12} - \rho_{13}\rho_{23}}{1 - \rho_{23}^2}\right), \qquad b_{21}^\star = \frac{\sigma_2}{\sigma_1}\left(\frac{\rho_{12} - \rho_{13}\rho_{23}}{1 - \rho_{13}^2}\right) \tag{22}$$

where the $[.]^\star$ denotes optimal (MMSE) estimates.

Uses of partial correlation

- In time-series analysis, partial correlations are used in devising what are known as partial auto-correlation functions, which are used in determining the order of auto-regressive models
- The partial cross-correlation function and its frequency-domain counterpart, known as partial coherency function, find good use in time-delay estimation
- In model-based control applications, partial correlations are used in quantifying the impact of model-plant mismatch in model-predictive control applications.

Semi-partial correlation: Discounting selectively

Semi-partial correlation

It examines the correlation between X_1 and **residualized** X_2 (w.r.t. X_3)

$$\rho_{1(2.3)} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{1-\rho_{23}^2}}$$

- "Closer" to regression coefficient than the partial correlation (why?)
- Smaller in value than the partial correlation coefficient (why?)

Commands in R

Commands	Utility
`mean,var,sd`	sample mean, variance and standard deviation
`colMeans,rowMeans`	means of columns and rows
`median,mad`	sample median and median absolute deviation
`cov,corr,cov2cor`	covariance, correlation and covariance-to-correlation
`lm, summary`	linear regression and summary of fit

Computing partial correlation in R

Use the `ppcor` package (due to Seongho Kim)

- `pcor`: Computes partial correlation for each pair of variables given others.

  Syntax: `pcormat <- pcor(X)`

  where $\mathbf{X}$ is a matrix with variables in columns. The default is Pearson correlation, but it can also compute Kendall and Spearman (partial) correlations. Result is given in `pcormat$estimate`

Computing semi-partial correlation in R

Use the `ppcor` package (due to Seongho Kim)

- `spcor`: Computes **semi-partial correlation**

  Syntax: `spcormat <- spcor(X)`

  The semi-partial correlation between X and Y given Z is computed as the correlation between X and $Y.Z$, i.e., only Y is conditioned on Z. The matrix of semi-partial correlations is *asymmetric*

Sample usage

```
w <- rnorm(200); v <- rnorm(200); z <- rnorm(200);      # ...
    Generate w,v and z
x <- 2*z + 3*w; y = z+ v;         # Generate x and y
cor(cbind(x,y))          # Compute correlation matrix between x ...
    and y
pcor(cbind(x,y,z))$estimate     # Compute partial correlation ...
    matrix
```